

Innovating Accelerated Use of NOAA Satellite Data – The Development of Accelerator-based Models and Applications

Allen Huang (allenh@ssec.wisc.edu), Bormin Huang, Jarno Mielikainen, Melin Huang

Space Science and Engineering Center, University of Wisconsin–Madison



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

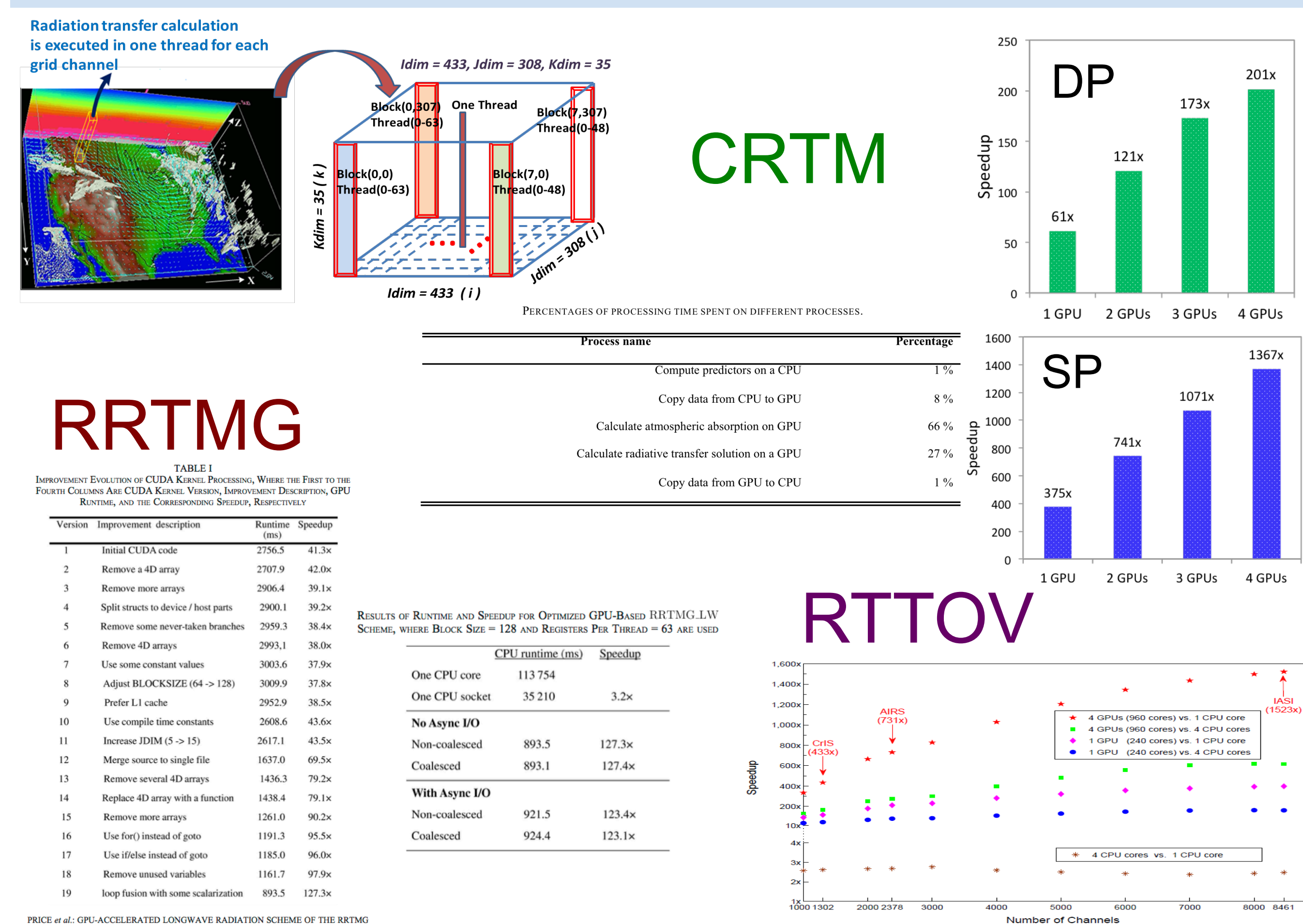


Introduction

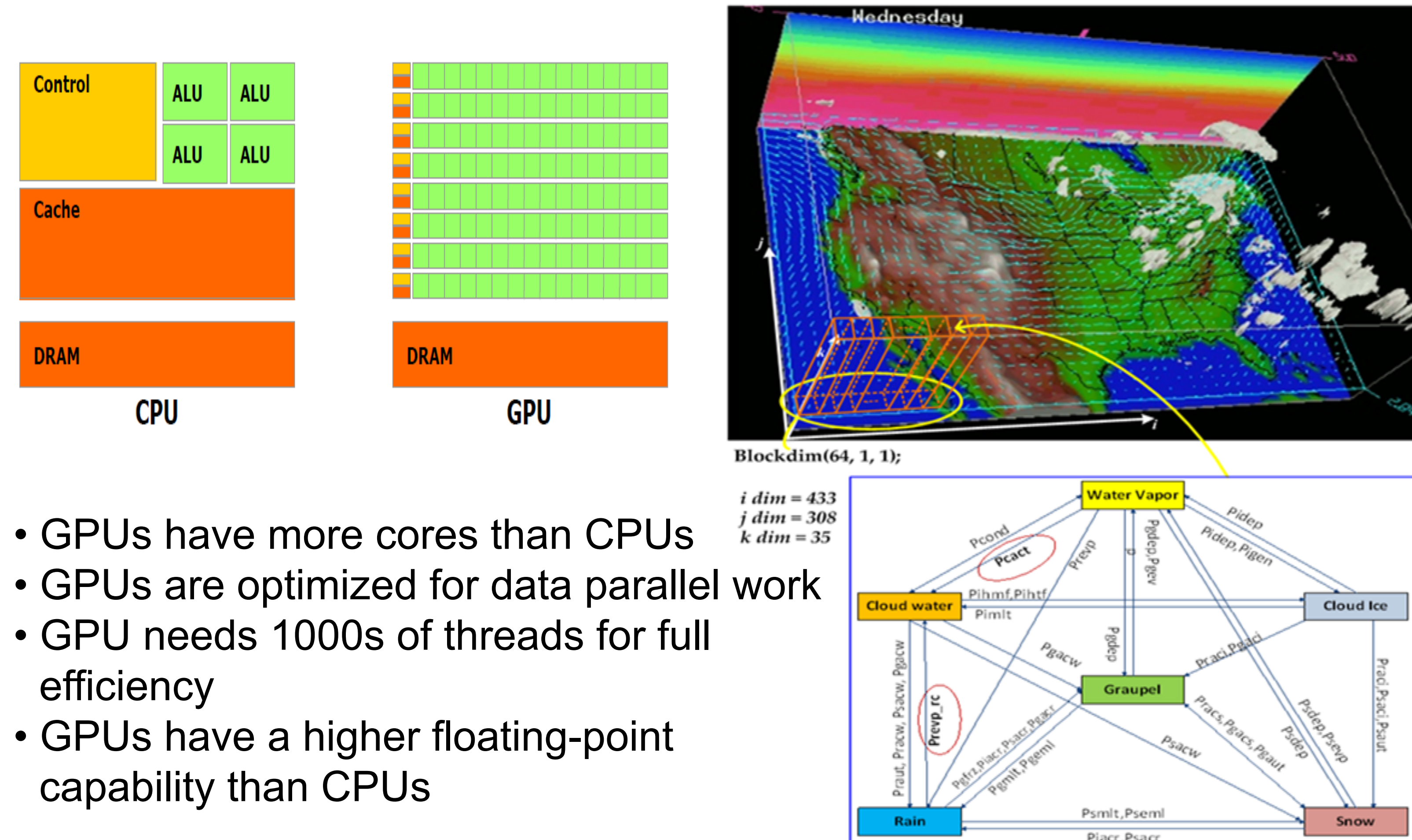
In the past 5 years scientists and engineers at SSEC has devoted their ingenuity to leveraging high-performance “accelerator” technology of NVIDIA Graphic Processing Unit (GPU) and more recently, the Intel Many Integrated Core (MIC) to advance satellite and weather forecasting applications. In 2014, this SSEC team, located at University of Wisconsin-Madison, was selected as one of the Intel Parallel Computing Center (IPCC). In 2015, NVIDIA awarded SSEC to develop a hybrid CPU-Fortran and GPU-CUDA weather forecasting model to demonstrate the impact of a highly accelerated radiative transfer model to pave the way for the timely, frequent, and optimal use of large volume of current and next generation of NOAA Low Earth Orbit (LEO) and Geosynchronous Equatorial Orbit (GEO).

In this poster we review the successful implementation of GPU-based high-performance radiative transfer models, such as CRTM, RTTOV, and RRTMG, running on NVIDIA GPUs via CUDA (Compute Unified Device Architecture). We continue with a review of the progress made so far in the development of a GPU based high-performance Weather Research Forecasting (WRF) model and demonstrate the design of a complete end-to-end GPU-CUDA WRF version, which could deliver a performance estimated to be >>10X speedup with respect to a single, modern CPU core.

GPU Speedups – Radiative Transfer Models (CRTM/RTTMG/RTTOV)



Parallel Execution of WRF on GPU



- **CPUs** focus on per-core performance - **Sandy Bridge**: 4 cores, **115.2 GFLOPS**, 90 W (~1.4 GFLOPS / W), Memory Bandwidth: **34.1 GB/s**
- **GPUs** focus on parallel execution - **Tesla K40**: 2,880 cores, **4290 GFLOPS**, 235 W (~18.3 GFLOPS / W), Memory bandwidth: **288 GB/s**

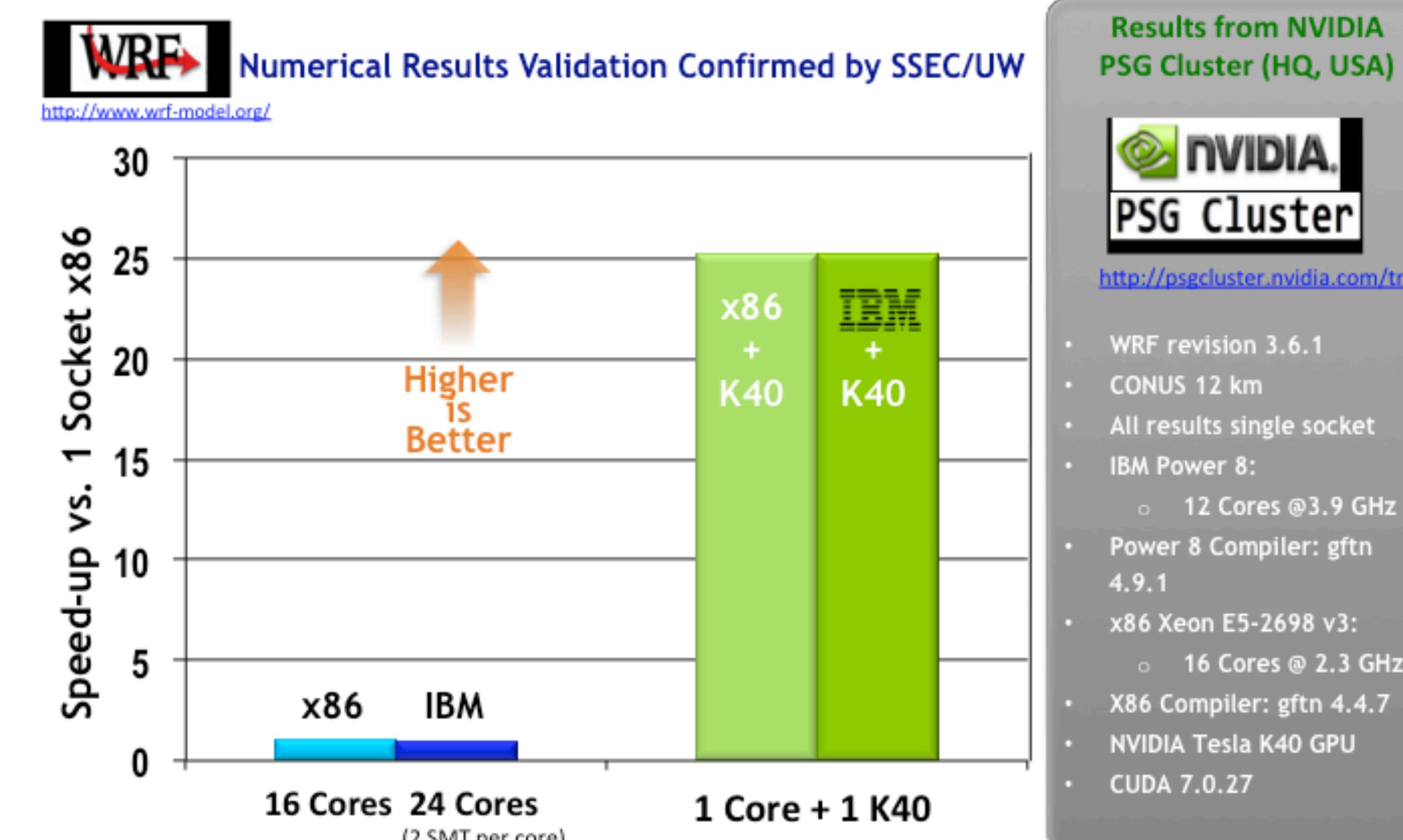
GPU Speedups – WRF Modules

Single threaded non-vectorized CPU code is compiled with gfortran 4.4.6

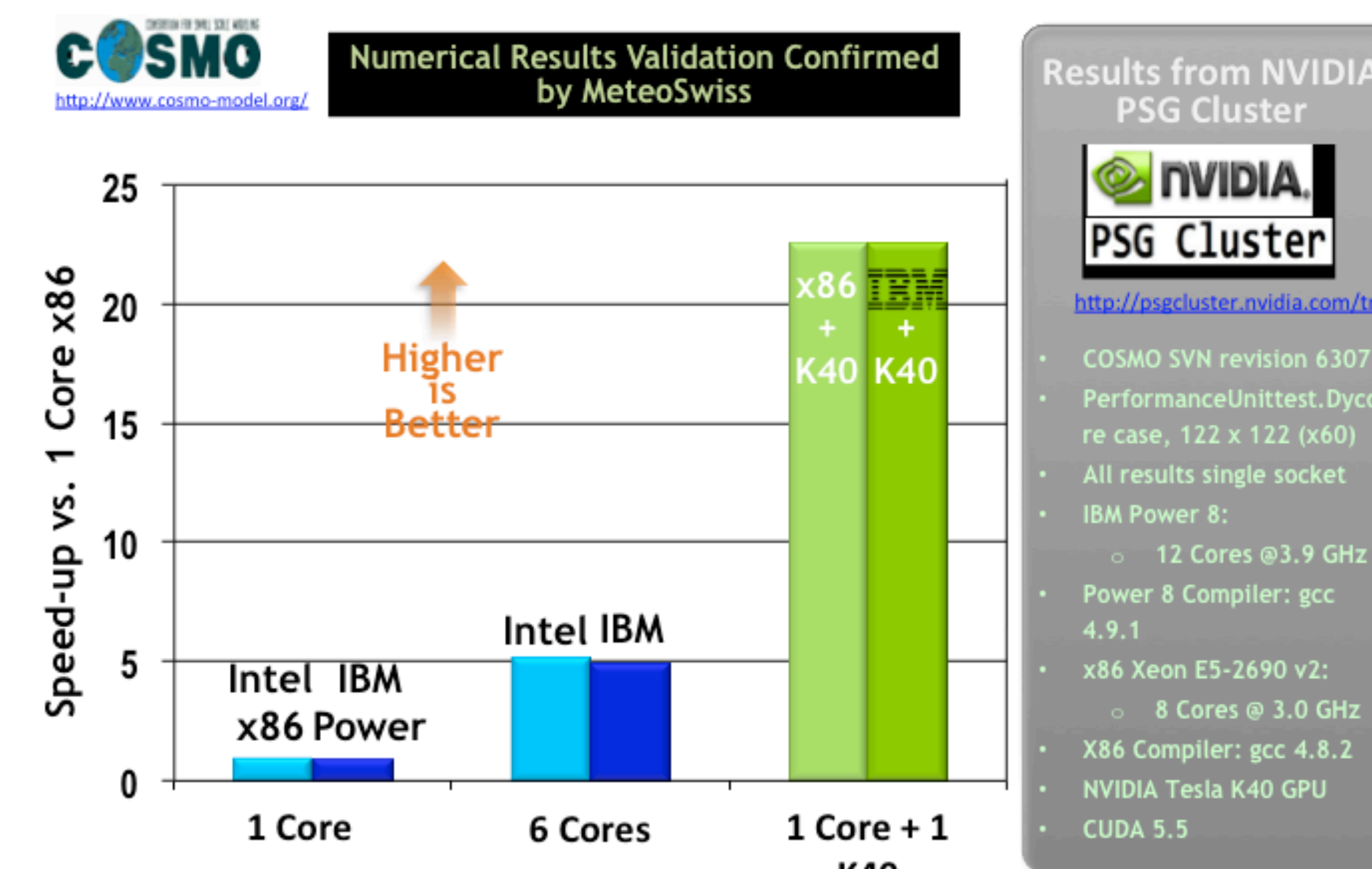
WRF Module name	Speedup
Single moment 6-class microphysics	500x
Eta microphysics	272x
Purdue Lin microphysics	692x
Stony-Brook University 5-class microphysics	896x
Betts-Miller-Janjic convection	105x
Kessler microphysics	816x
New Goddard shortwave radiance	134x
Single moment 3-class microphysics	331x
New Thompson microphysics	153x
Double moment 6-class microphysics	206x
Dudhia shortwave radiance	409x
Goddard microphysics	1311x
Double moment 5-class microphysics	206x
Total Energy Mass Flux surface layer	214x
Mellor-Yamada Nakanishi Niino surface layer	113x
Single moment 5-class microphysics	350x
Pleim-Xiu surface layer	665x

Performance Comparison

WRF WSM6 Results for Power8 + K40



COSMO Dynamics Results for Power8 + K40



Summary

- **Intel** awarded SSEC a two-year grant to develop Intel MIC Xeon Phi Coprocessor based WRF using open ACC common architecture
 - SSEC becomes one of the Intel Parallel Computing Center (IPCC)
- **NVIDIA**, world largest GPU chip maker, is funding SSEC to develop a GPU-CPU Hybrid WRF prototype using CUDA architecture
- **IBM** Power architect, WRF experiment with POWER CPU+ NVIDIA GPU
- **TQI**, GPU/CUDA based WRF for Low Latency Wx Forecast applications